DEPARTMENT OF THE ARMY
U.S. Army Corps of Engineers
Washington, DC 20314-1000

CECW-EH

ETL 1110-2-537

Technical Letter
No. 1110-2-537

31 October 1997

Engineering and Design
UNCERTAINTY ESTIMATES FOR
NONANALYTIC FREQUENCY CURVES

Distribution Restriction Statement

Approved for public release; distribution is unlimited.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
31 Oct 1997	N/A	-

Title and Subtitle Engineering and Design: Uncertainty Estimates for Nonanalytic Frequency Curves	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Department of the Army U.S. Army Corps of Engineers Washington, DC 20314-1000	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes	
Abstract	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU
Number of Pages 11	

CECW-EH

DEPARTMENT OF THE ARMY
U.S. Army Corps of Engineers
Washington, DC 20314-1000

ETL 1110-2-537

Technical Letter
No. 1110-2-537

31 October 1997

Engineering and Design
UNCERTAINTY ESTIMATES FOR NONANALYTIC FREQUENCY CURVES

1-1. Purpose

This letter presents a methodology for computing the uncertainty about nonanalytic frequency curves. This situation arises when estimating flow-frequency curves from hypothetical events, regulated frequency curves, and stage-frequency curves. The method involves the application of order statistics to compute the uncertainty distribution.

1-2. Applicability

This letter applies to all HQUSACE elements and USACE commands where estimates of uncertainty about nonanalytic frequency curves are required. The primary expected application is in developing uncertainty relationships for use in risk-based analysis of flood damage reduction projects.

1-3. References

a. EM 1110-2-1619. Risk-Based Analysis for Flood Damage Reduction Studies.

b. Dixon, W. J., and Massey, F. J., Jr. (1957). *Introduction to Statistical Analysis*. McGraw Hill, New York, 488 pp.

c. Kottegoda, N. T. (1980). *Stochastic Water Resources Technology*. Halsted Press, New York, 384 pp.

d. Mood, M. A., Graybill, F. A., and Boes, D. C. (1963). *Introduction to the Theory of Statistics*. McGraw-Hill, New York, 564 pp.

e. Press, W. H., Flannery, B. P. , Teukosky, S. A., and Vettering, W. T. (1990). *Numerical Recipes (Fortran)*. Cambridge University Press, New York, 702 pp.

1-4. Distribution Statement

Approved for public release, distribution is unlimited.

1-5. Uncertainty Estimation

a. General.

(1) The uncertainty in a frequency curve that is estimated from a graphical fit of ordered observations (e.g., peak annual regulated flows or stages) may be calculated from order statistics. No assumption need be made concerning the analytic form of the frequency curve. Under these circumstances, the statistic derived to estimate uncertainty is termed “nonparametric” or, more to the point, “distribution free.” Note that the procedures outlined in EM 1110-2-1619, Section 4, should be utilized if an analytic distribution such as the log-Pearson III can be used to approximate the frequency curve.

(2) The order statistic approach is limited to calculating uncertainty in the estimated frequency curve for the range of observed data or, alternatively, the equivalent length of record. Extrapolating the estimates beyond the range of data is performed by using asymptotic approximations of uncertainty distributions. The order statistic and asymptotic estimates of uncertainty are matched at the limits of the observed data.

b. Order statistics.

(1) The order statistic approach relies on a very straightforward application of the binomial distribution (Mood, Graybill, and Boes 1963, p 513). The problem is to calculate the probability that the flow or stage corresponding to an exceedance probability exceeds a particular value. The only values available are the observations of the random variable. Assume that the observations of flow or stage are ordered as Y_j , j = 1,2,3,...n, $Y_j \leq Y_{j+1}$ and n is the number of years of record. The uncertainty about the frequency curve at nonexceedance probability p is estimated by the probability

$$P[\,Y_p \geq Y_j\,] \;=\; P[\text{ jth order statistic } \leq \; Y_p\,]$$
$$= P[\text{ j or more observations } \geq Y_p\,]$$

where Y_p = quantile (e.g., flow or stage) for nonexceedance probability p. Applying the binomial theorem

$$P[\,Y_p \geq Y_j\,] \;=\; \sum_{i=j}^{n} \binom{n}{i} \,[\,F(Y_p)\,]^i\,[\,1 - F(Y_p)\,]^{n-i} \tag{1}$$

where $p = F\,(\,Y_p\,)$ = nonexceedance probability associated with the quantile of interest. The nonexceedance probability or a corresponding exceedance probability can be calculated by use of a plotting position formula.

(2) This computation is rather inconvenient, although it can be performed fairly easily by computer. A more convenient and equivalent expression which estimates the uncertainty in the estimate of the nonexceedance probability p for quantile Y_p , involves the incomplete beta function (Press et al. 1990, p 166)

$$P[F(Y_p) \geq F(Y_j)] \;=\; \frac{1}{B(j, n-j+1)} \int_0^p z^{j-1}(1-z)^{n-j}\,dz \tag{2}$$

$$= IB_p(j, n-j+1) \tag{3}$$

where

$F\,(Y_p\,) = p\;B\,(\,j,\,n - j\,+\,1\,)$ = beta function

$IB_{\,p}\,(\,j,\,n - j\,+\,1\,)$ = incomplete beta function

This expression gives the probability that the nonexceedance probability associated with Y_p is greater than or equal to the nonexceedance probability associated with the jth ordered observation $F\,(\,Y_j\,)$. The equivalent relationship for exceedance probabilities is

$$P[\,F_e(Y_p) \;\leq\; F_e(Y_j)\,] \;=\; IB_p(\,j, n-j+1) \tag{4}$$

where

$F_e\,(\,Y_j\,)$ = exceedance probability associated with the jth observed data point

$F_e\,(\,Y_p\,) = (\,1 - p\,)$ = exceedance probability associated with Y_p

(3) The important characteristic of these equations is that the distribution estimate $P\,[F_e\,(\,Y_p\,) < F_e\,(\,Y_j\,)]$, or confidence level, is associated with a rank j, and consequently a ranked observation Y_j . Therefore, estimates of uncertainty can only be provided corresponding to the range of the Y_j . This is shown in Figure 1. Note that the uncertainty distribution cannot be extrapolated beyond the smallest or largest observation. If only 20 observations exist, then the maximum value for any uncertainty distribution level is the largest out of 20 observations.

(4) Despite this limitation, the order statistic estimate is useful because it is sensitive to local changes of slope in the frequency curve. This means the method is able to account for the local variability of a random variable. For example, a regulated frequency curve may have very small slope over a significant range in probability indicating very little variability in releases. The order statistic approach will recognize this and predict very little uncertainty in the estimated frequency curve over this range of probabilities.

c. Example. As a simple example of the application of Equation 1, assume that there are 5 years of observed peak annual stages. The estimated nonexceedance probability corresponding to the largest observed event is computed by the Weibull plotting position as

$$F(Y_p) = m/(n+1) = 0.833$$

where

m = event rank

n = number of years of record

Given that the 0.833 nonexceedance probability event or 0.167 exceedance probability event occurs, the probability that the corresponding quantile Y_p exceeds the third-ranked observation, $Y_j = Y_3$ can be computed using Equation 1:

i (1)	$^*c$ (2)	$F(Y_p)^i$ (3)	$(1-F(Y_p))^i$ (4)	$^{**}p$ (5)
3	5!/(3!2!)=10.0	0.833^3	0.167^2	0.159
4	5!/(4!1!)=5.0	0.833^4	0.167^1	0.396
				0.410
5	5!/(5!0!)=1.0	0.833^5	0.167^0	0.96

*Binomial coefficient.
**Product of columns 2,3,4.

Consequently, there is a 0.96 probability, or 96 percent chance, that the stage corresponding to the 0.167-exceedance probability Y_p exceeds the third-ranked event Y_3.

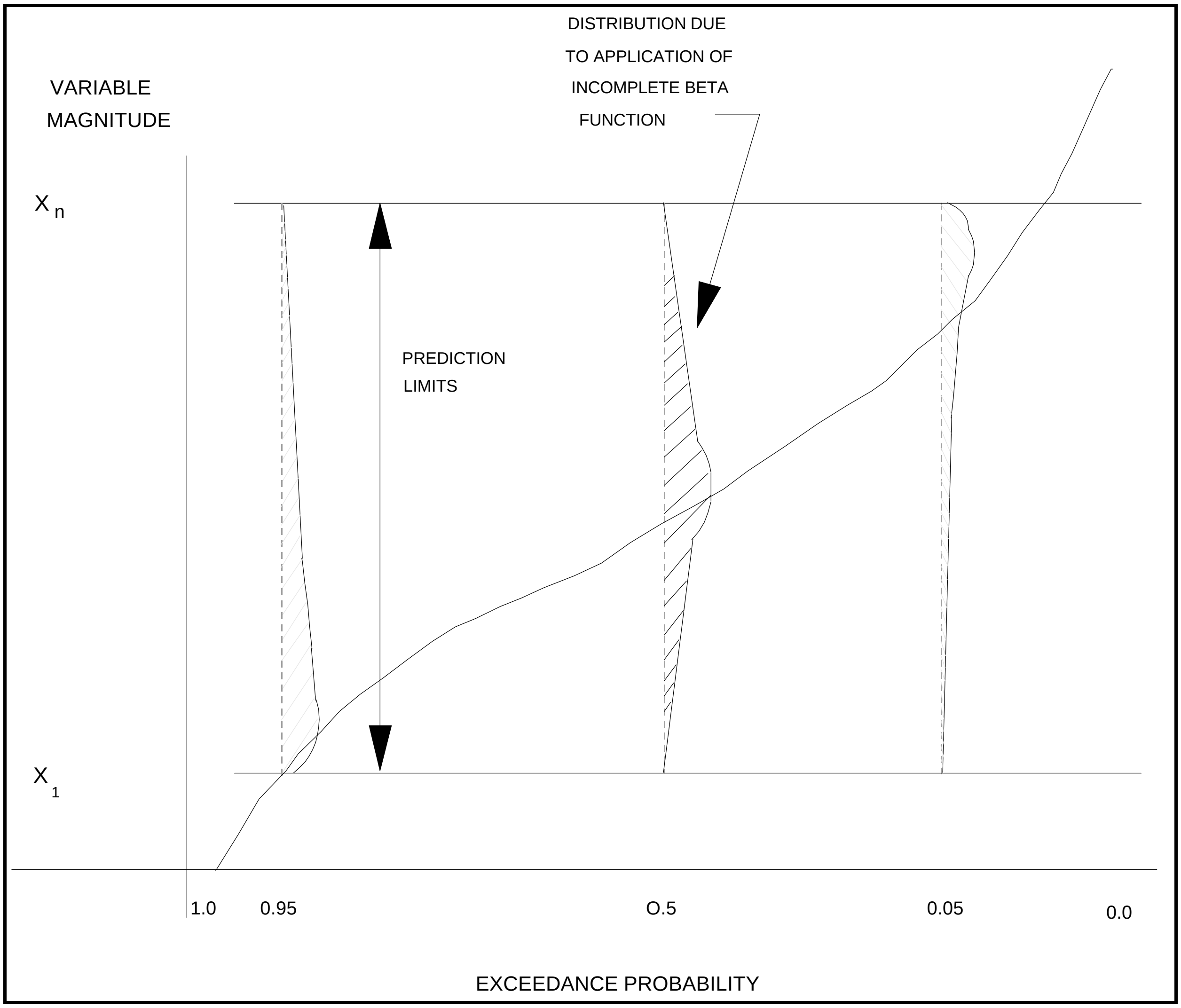

Figure 1. Uncertainty with order statistics

d. Equivalent years of record. The discussion refers to estimating the uncertainty for ordered observations. However, the theory can be extended if an equivalent years of record is available for an estimated frequency curve. For example, assume that the information used to develop a flow-frequency curve from hypothetical events is worth 20 years of equivalent record length. Plotting positions for the 20 years can be calculated and the corresponding Y_j determined from the hypothetical frequency curve. The Y_j can then be used as previously described to calculate confidence limits.

e. Asymptotic approximations.

(1) The asymptotic approximation used for the uncertainty distribution differs depending on whether or not stage or flow is the focus. In the case of stage, the uncertainty is determined based on the result that the binomial distribution is asymptotically normal for large n. Under this assumption, the variance of a normal uncertainty distribution for quantile X is given as

$$S_X^2 = \frac{p(1-p)}{n f_X(x)^2} \tag{5}$$

where

S_X = standard deviation of the uncertainty distribution

p = exceedance probability

n = record length

$f_X(x)$ = probability density function for the frequency curve of interest

Note that the probability density function is the inverse of the slope of the nonexceedance frequency curve.

(2) A suggested guideline for applying this asymptotic approximation given by Dixon and Massey (1957) is that $np \geq 5$. Given this criterion, the method could be applied at most to the 0.25-nonexceedance probability or about a 2-year event for n = 20 years. Given this limited range, the order statistic approach is used to estimate uncertainty for the range of the data, and the asymptotic approximation is used to extrapolate the results beyond the range of data.

(3) The above approximation is particularly useful for extrapolating the frequency curve for stages corresponding to the channel floodplain elevation. Under these circumstances, the frequency curve is relatively flat and the sampling uncertainty in the frequency curve actually decreases with increasing deviation from the mean stage. The above approximation captures this effect because the standard deviation is inversely proportional to the local value of the probability density function. The behavior of the standard deviation in this circumstance is very different than that for unregulated streamflow frequency curves where the standard deviation increases with distance from the mean flow.

(4) Although general, the above approximation is too conservative for application to hypothetical and regulated frequency curves which are approximately analytic for extreme probabilities, or for stage-frequency curves which do not have broad overbank areas. Consequently, as an alternative approach, the order statistic approximation of uncertainty is extended for these flow-frequency curves using the following asymptotic approximation to the uncertainty in a normal distribution (Kottegoda 1980, p 233)

$$S_x^2 = \frac{S^2}{n} + Z_p^2 \frac{S^2}{2n} \tag{6}$$

where

S = standard deviation of frequency curve

Z_p = normalized deviate computed as

$$Z_p = \frac{X - M}{S} \tag{7}$$

where

M = mean of the frequency curve

X = flow corresponding to exceedance probability p

Equation 6 is used exclusively for hypothetical or regulated frequency curves and is used to constrain the variance estimated by Equation 5.

f. Plotting position. Numerous plotting position formulas have been proposed in the past. However, estimating the expected annual damage presumes that the expected value of the frequency curve (the expected probability estimate) is preserved in the process. Consequently, the Weibull plotting position is preferred, because it is the expected estimate of the exceedance or nonexceedance probability.

1-6. Application to Estimated Frequency Curves

a. Methodology. The order statistic approach provides estimates of uncertainty for a limited range. An approach is suggested herein for utilizing the order statistic estimates to obtain the uncertainty distribution for the range of frequency curve probabilities of interest. The approach taken is to fit a normal distribution to the confidence limits obtained with the order statistic approach. The normal distribution was selected because it matches the asymptotic approximation used to extrapolate uncertainty estimates. Furthermore, the normal distribution is convenient for use with a Monte Carlo simulation. The general steps involved in estimating the equivalent normal distribution are as follows:

(1) Compute the order statistic confidence limits for the computed plotting positions corresponding to the appropriate record length. A natural selection of the location and number of points would be at the ordered observations.

(2) Calculate the mean and standard deviation of the uncertainty distributions developed from the incomplete beta function at each of the points selected in step (1). The calculation is approximate because a full range of probabilities for the uncertainty distribution is not obtained from the order statistic approach. The calculation of uncertainty is limited by the range of observations or record length as was described previously (Figure 1). Consequently, the mean and standard deviation are computed only for uncertainty distributions with exceedance probabilities defined minimally between 0.9 and 0.1. The mean and standard deviation for the uncertainty distribution computed with Equation 2 are computed based on a probability weighted basis using trapezoidal rule integration as

$$M = \sum_{i=2}^{n} 0.5\,(Y_{(i-1)} + Y_i)\ \frac{IB_p(i, n-i+1) - IB_p(i-1, n-i))}{IB_p(n,1) - IB_p(1,n)} \qquad (8)$$

$$S^2 = \sum_{i=2}^{n} 0.5((Y_{i-1} - M)^2 + (Y_i - M)^2)\frac{IB_p(i,n-i+1) - IB_p(i-1,n-i)}{IB_p(n,1) - IB(1,n)} \qquad (9)$$

where M and S are estimates of the sample mean and standard deviation of the uncertainty distribution computed for any quantile using the incomplete beta function.

(3) The standard deviation computed in step (2) is set equal to that of the normal distribution to compute uncertainty in the frequency curve. In other words, the standard deviation of the distributions computed with the incomplete beta function is set equal to that of the normal distribution to obtain an approximate uncertainty distribution. A limitation might be placed on the use of the standard deviations computed in the previous step in estimating the equivalent normal distribution. In general, a reasonable expectation is that the uncertainty should be nondecreasing once a maximum value has been reached. Two of these maxima will occur near the extreme ends of the plotted points. Consequently, the maximum values of the standard deviation computed should define the range where the normal distribution standard deviation is equated to that computed in step (2). Let Y_m and Y_l be the ordered observations where the corresponding maximum variances S_m and S_l have been determined.

(4) The calculation of the normal approximation in step (3) is only useful for computing the uncertainty distribution for the range of quantiles (e.g., stages or flows) used in step (2). The asymptotic approximations for either stage, Equation 5, or hypothetical or regulated frequency curves, Equation 6, are matched at this point to the order statistic estimates by equating variances at Y_m and Y_l. For example, consider equating variance for a stage-frequency curve. This is done by solving Equation 5 for an equivalent record length for extrapolation. For quantile values greater than Y_m this becomes

$$n_m = \frac{p(1-p)}{S_m^2 f_Y(Y_m)^2} \qquad (10)$$

where

p = exceedance probability for Y_m

$f_Y(Y_m)^2$ = probability density function

The inverse of the slope of the estimated nonexceedance frequency curve is used to compute the density function. Correspondingly, the order statistic and asymptotic approximation estimates are matched for quantiles less than Y_l as

$$n_l = \frac{p(1-p)}{S_l^2 f_Y(Y_l)^2} \tag{11}$$

Equation 5 can now be used to extrapolate the estimates of uncertainty using the computed values of n_m and n_l . A similar procedure is used to obtain equivalent record lengths from Equation 6 for hypothetical or regulated frequency curves.

b. Scenarios.

(1) Estimation of frequency curves involves various kinds of information. Ideally, a sufficient number of gauged observations are available to estimate the frequency curve. This is rarely the case. If gauged observations are not available, rainfall-runoff analysis can be used to develop a synthetic frequency curve. The equivalent years of record are determined for this frequency curve from Table 1 reproduced here from Table 4-5 of EM 1110-2-1619. The order statistic methodology can be used to estimate the uncertainty in the frequency curve given the equivalent years of record as described in paragraph 1-5*d*.

(2) In some cases, a mixture of gauged observations and hypothetical events is used to estimate the frequency curve. The gauge record length is only sufficient to estimate the frequency curve for relatively high-frequency events. The hypothetical events are used to extend the frequency curve. The confidence and corresponding uncertainty in the frequency curve differs for the region of the curve defined by the observed or hypothetical events.

(3) In this case, the uncertainty in the region of the frequency curve determined by the observed events is calculated using the observed record length. Likewise, the uncertainty in the region influenced by the hypothetical events is determined by an appropriate equivalent years of record. A transition region exists where the frequency curve is defined between the largest observed event and the smallest hypothetical event by interpolation. The recommendation herein is to determine the uncertainty in this region based on a weighted average of the uncertainty calculated for the largest observed event and smallest hypothetical event. The weight is inversely proportional to the distance from the respective events plotted on probability paper.

c. Limitations. Extrapolation of a frequency curve needs to be constrained by the characteristics of the field conditions. This is also true of uncertainty distributions. Typical situations where the field conditions need to be considered are in applications to regulated frequency curves. Regulated frequency curves should approach the inflow frequency curves as the effect of the reservoir diminishes. Certainly, the uncertainty distributions should also become equivalent as these two frequency curves become equivalent.

d. Example 1. The estimation of the uncertainty distribution for a stage-frequency curve on the Sacramento River will be used to demonstrate the procedure outlined previously. The frequency curve developed for the location is shown in Figure 2 from data listed in Table 2. Also shown in Table 2 is the mean and standard deviation for the frequency curve which was computed using trapezoidal rule integration.
If one assumes that the frequency curve was estimated using information that is worth 20 years of record, ordered values are obtained from the frequency curve using the Weibull plotting position as shown in Table 3. The uncertainty distribution will be computed by following the steps outlined in paragraph 1-6*a*.

(1) Step 1: The uncertainty distribution for each ordered value was computed using the incomplete beta function via Equation 2.

(2) Step 2: The mean and standard deviation of the uncertainty distribution at each ordered value was computed via Equations 6 and 7 and are shown in Table 4. Note again that these moments are only computed for uncertainty distributions that can be defined for exceedance probabilities minimally ranging from 0.1 to 0.9.

(3) Step 3: Inspection of Table 4 indicates that the maximum standard deviations are S_l = 2.35 and S_m = 1.54 corresponding to stages Y_l = 12.33 and Y_m = 21.29. The standard deviation of the normal distribution approximation to the uncertainty distribution

Table 1
(ref: Table 4-5, EM 1110-2-1619) Equivalent Record Length Guidelines*

Analysis Setting	Equivalent Record Length
1. Long-period gauged record available at site	Use systematic record length
2. Long-period gauge on stream, drainage area within 20%	Use 90% to 100% of record length
3. Long-period gauge within watershed, model calibrated to gauge-based curve	Use 50% to 90% of record length
4. Calibrate to regional frequency curve	Average years from regional study
5. Short-period event gauge in watershed, model calibrated to several events	Use 20 to 30 years record length
6. Regional model parameter data available, no calibration	Use 10 to 30 years record length
7. Handbook/textbook model coefficients only	Use 10 to 15 years record length

*Judgment should be applied to account for the quality of the data used in the analysis, degree of confidence in validated model, and previous experience in similar studies in the area.

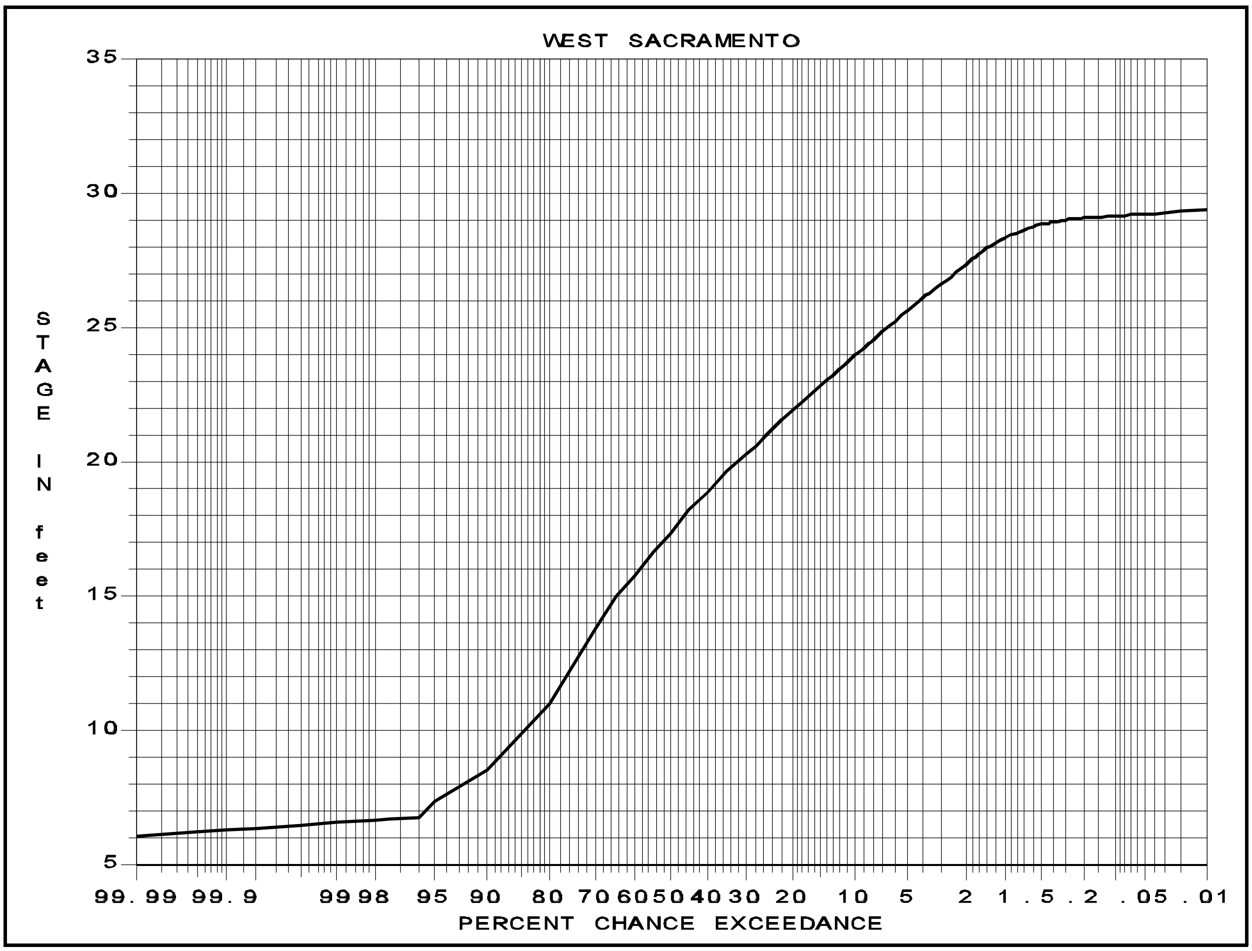

Figure 2. Example stage-frequency curve

within the range of values defined by Y_l and Y_m is determined by the values in Table 4.

Table 2
Example Stage-Frequency Curve

Event Order	Stage	Exceedance Frequency
1	6.60	0.9900
2	6.80	0.9600
3	7.40	0.9500
4	7.95	0.9250
5	8.55	0.9000
6	8.95	0.8600
7	9.95	0.8400
8	11.05	0.8000
9	12.70	0.7500
10	13.85	0.7000
11	14.90	0.6600
12	15.80	0.6000
13	16.70	0.5500
14	17.40	0.5000
15	18.25	0.4500
16	18.90	0.4000
17	19.70	0.3500
18	20.30	0.3000
19	21.10	0.2500
20	21.95	0.2000
21	24.00	0.1000
22	25.70	0.0500
23	27.40	0.0200
24	27.60	0.0180
25	27.80	0.0160
26	28.00	0.0140
27	28.20	0.0120
28	28.40	0.0100
29	28.90	0.0050
30	29.10	0.0025

Note: Frequency curve integral moments:
mean = 17.00 ft, std dev = 5.60 ft.

Table 3
Plotting Positions Estimated from Stage-Frequency Curve

Years of Record	Stage	Probability
1	7.27	0.9524
2	8.45	0.9048
3	9.10	0.8571
4	10.80	0.8095
5	12.33	0.7619
6	13.53	0.7143
7	14.73	0.6667
8	15.52	0.6190
9	16.32	0.5714
10	17.07	0.5238
11	17.80	0.4762
12	18.53	0.4286
13	19.20	0.3810
14	19.90	0.3333
15	20.52	0.2857
16	21.29	0.2381
17	22.11	0.1905
18	23.00	0.1429
19	24.13	0.0952
20	25.80	0.0476

(4) Step 4: Calculation of n_m and n_l is performed using Equations 8 and 9. Application of the equations requires that an estimate of the probability density function be obtained at each point as the inverse of the slope of the nonexceedance frequency curve. From Table 4, the data are obtained to take a centered difference approximation to this slope. The inverse slope at Y_m = 21.29 is

$$f_Y(Y_l) \sim \frac{0.810 - 0.714}{13.53 - 10.80} = 0.035 \qquad (12)$$

and at Y_l = 12.33 is

$$f_Y(Y_m) \sim \frac{0.333 - 0.238}{22.11 - 20.52} = 0.0498 \tag{13}$$

Table 4
Mean and Standard Deviations of Uncertainty Distributions Computed Via Order Statistics

Frequency Curve		Statistics of Uncertainty from Incomplete Beta	
Probability	Quantile	Mean	Std Dev
0.905	8.45	9.15	1.52
0.857	9.10	10.13	1.96
0.810	10.80	11.26	2.26
Beginning of data used in interpolation			
0.762	12.33	12.39	2.35
0.714	13.53	13.47	2.31
0.667	14.73	14.48	2.16
0.619	15.52	15.40	1.98
0.571	16.32	16.21	1.87
0.524	17.07	16.98	1.75
0.476	17.80	17.71	1.66
0.429	18.53	18.40	1.60
0.381	19.20	19.07	1.53
0.333	19.90	19.76	1.52
0.286	20.52	20.46	1.53
0.238	21.29	21.18	1.54
End of data used in interpolation			
0.190	22.11	21.93	1.53
0.143	23.00	22.67	1.48
0.095	24.13	23.48	1.38

The equivalent record lengths are obtained using S_m = 1.54 and p = 0.286 in Equation 8

$$n_m \sim \frac{0.286(1 - 0.286)}{(1.53)^2(0.0498)^2} \sim 35 \tag{14}$$

and S_l = 2.35 and p = 0.762 in Equation 9

$$n_m \sim \frac{0.762(1 - 0.762)}{2.35^2(0.035)^2} \sim 27 \tag{15}$$

A straightforward application of the normal distribution can now be used to extrapolate the uncertainty distribution. The resulting confidence limits at plus or minus one and two standard deviations associated with the uncertainty calculation are shown in Figure 3 for the entire frequency curve.

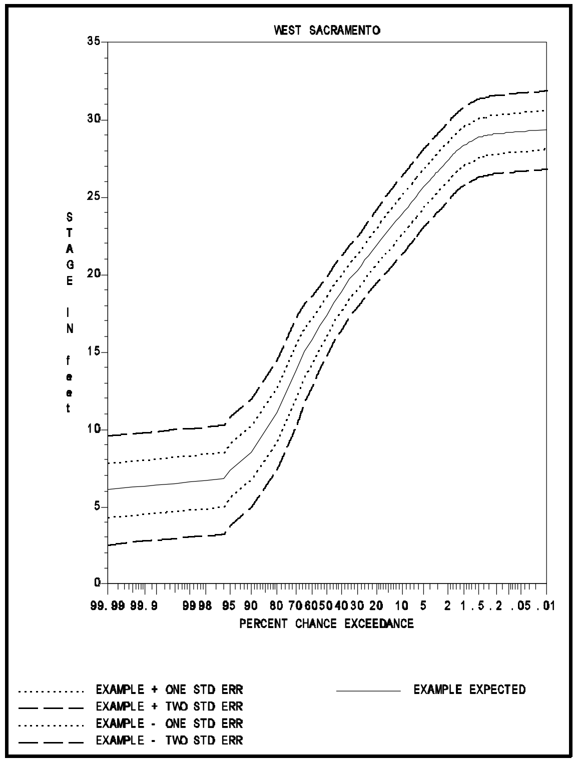

Figure 3. Stage-frequency curve with confidence limits

e. Example 2.

(1) The estimated uncertainty distribution for a regulated frequency curve on the Savannah River was calculated, resulting in the confidence limits shown in Figure 4. The frequency curve was based on information from a long inflow series, historic information, observed regulated flows, and model simulations. The equivalent record length from this information was believed to be about 190 years.

(2) The interesting aspect of this example is that the uncertainty about the frequency curve is negligible as the frequency curve flattens near the 50 percent chance exceedance probability event. This is indicated by the convergence of the confidence limits in Figure 4.

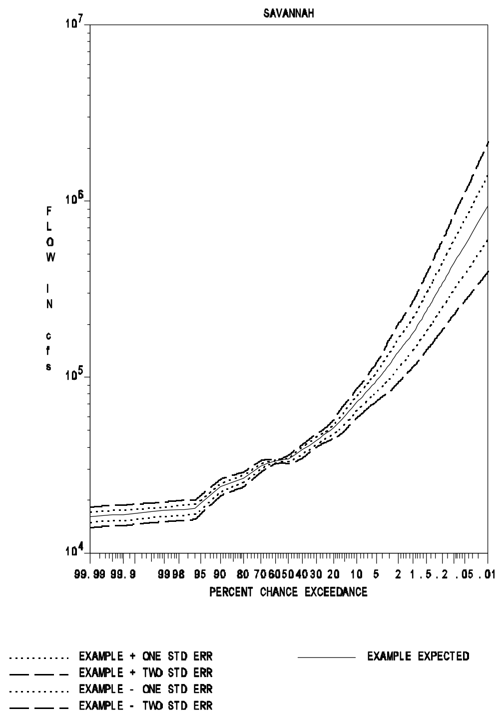

Figure 4: Regulated-frequency curve with confidence limits

Figure 4. Regulated frequency curve with confidence limits

The example clearly demonstrates the capability of the order statistic method to sense the local changes in the potential variability of flow values.

FOR THE COMMANDER:

STEVEN L. STOCKTON, P.E.
Chief, Engineering Division
Directorate of Civil Works